Association Rules

Rita P. Ribeiro

Machine Learning - 2021/2022

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

[dcc] DEPARTAMENTO DE CIÊNCIA DE COMPUTADORES
FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO

Summary

Association Rules in Action

Association Rules: a New Data Mining Task

Data Mining Tasks:

- Predictive
  - Classification
  - Regression
  - ...
- Descriptive
  - Clustering
  - Association Rules
    - find relationships / associations between groups of variables
  - ...

Motivation

Originally stated in the context of Market Basket Analysis

- Data consists of set of items bought by costumers, referred as transactions
- Find unexpected associations between sets of items using the frequency of sets of items
- Discovered sets of items are referred as frequent itemsets or frequent patterns
- Goals
  - Store layout - *Should products A and B be placed together?*
  - Promotions - *If the client is interested in {A,B,C,...}, can we guess other interests?*
  - ...

Actionable Knowledge: shop layout

- Possible actions from rule $\{A1, A4\} \rightarrow \{A6\}$
  - Sell the $A1$, $A4$, $A6$ together (pack)
  - Place article $A6$ next to articles $A1$, $A4$
  - Offer a discount coupon for $A6$ in articles $A1$, $A4$
  - Place a competitor of A6 next to $A1$, $A4$ (brand protection).
- Note
  - These actions must make sense from the business point of view.

Actionable Knowledge: cross selling

- Steps
  - Client puts article A in basket
  - Shop knows rule $A \rightarrow B$
  - Rule has enough confidence (> 20%)
  - Shop tells client he may be interested in B
  - Client decides whether to buy B or not
- Notes
  - Rules are discovered from business records
  - Discovery (mining) can be made off-line
  - Use of rules can be made on-line

Actionable Knowledge: text mining

- Each document is treated as a “bag” of terms and keywords
  - doc1: Student, Teach, School (Education)
  - doc2: Student, School (Education)
  - doc3: Teach, School, City, Game (Education)
  - doc4: Baseball, Basketball (Sport)
  - doc5: Basketball, Player, Spectator (Sport)
  - doc6: Baseball, Coach, Game, Team (Sport)
  - doc7: Basketball, Team, City, Game (Sport)
- Goal: identify co-occurring terms and keywords
- Example:
  - Student, School $\rightarrow$ Education
  - Game $\rightarrow$ Sport

Actionable Knowledge: health

- Rules obtained from the patient's records
- Sooner prevention
- Each patient visits a health unit one or more times
- We record the observations for each visit
    - Symptoms (head ache, temperature)
    - Exam results (blood pressure, sugar level)
- A set of observations may fire a rule
  {Head ache, blood pressure rise} $\rightarrow$ {stroke, immobilization}
- When head ache and blood pressure rise are observed, stroke and immobilization are also expected.
- Not necessarily causal

Actionable Knowledge: web usage analysis

Usage patterns

- Most visited pages
- Frequent page sets
  - Site structure
- Pages associated to users
  - personalization
- Seasonal effects
  - operations, campaigns
- Cross-preferences
  - cross-selling

Association Rules
Basic Concepts

Market Baskets data set

TID	Products
1	A, B, E
2	B, D
3	B, C
4	A, B, D
5	A, C
6	B, C
7	A, C
8	A, B, C, E
9	A, B, C

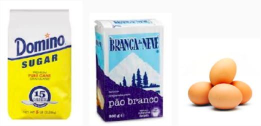
Products are converted in binary flags

→

TID	A	B	C	D	E
1	1	1	0	0	1
2	0	1	0	1	0
3	0	1	1	0	0
4	1	1	0	1	0
5	1	0	1	0	0
6	0	1	1	0	0
7	1	0	1	0	0
8	1	1	1	0	1
9	1	1	1	0	0

Market Basket Analysis: how frequent is an itemset?

- Sugar, Flower and Eggs are sold together



- How important is this set?

- **Support** measures the importance of a set
    - Percentage of transactions *t* containing the set *S*
    - Absolute support: number of transactions *t* containing the set *S*

Market Basket Analysis: how predictive is an itemset?

- Frequent itemsets are used to generate association rules.
- If you buy sugar and flower, you also buy eggs.
- How strong is this rule?
- **Confidence** measures the strength of the rule
  - Percentage of transactions *t* that having sugar and flower also have eggs

Association Rules: Basic Concepts

- Consider a set of items I
- A transaction t is a subset of items, i.e. $t \subseteq I$
- Given a data set of transactions $D = \{t_i\}_{i=1}^{N}$
- An association rule is defined as an implication $X \rightarrow Y$, where
  - X and Y are itemsets, i.e. $X, Y \subseteq I$
  - $X \neq \emptyset$, $Y \neq \emptyset$ and $X \cap Y = \emptyset$
- $sup(X)$ is the proportion of transactions in D that include the itemset X
- support: $sup(X \rightarrow Y) = sup(X \cup Y)$
- confidence: $conf(X \rightarrow Y) = sup(X \cup Y)/sup(X)$

Association Rules: an example

Given the data

Transactions ID	Items Bought
100	A, B, C
200	A, C
150	A, D
500	B, E, F

$\rightarrow$

TID	A	B	C	D	E	F
100	1	1	1	0	0	0
200	1	0	1	0	0	0
150	1	0	0	1	0	0
500	0	1	0	0	1	1

- The itemsets with a minimum support of 50%

Frequent Itemsets	Support
{A}	75%
{B}	50%
{C}	50%
{A,C}	50%

- Rules with minimum support of 50% and minimum confidence of 50%
  - $A \rightarrow C$
    - $sup(A \rightarrow C) = sup(\{A, C\}) = 50\%$
    - $conf(A \rightarrow C) = sup(\{A, C\})/sup(\{A\}) = 66.6\%$
  - $C \rightarrow A$
    - $sup(C \rightarrow A) = sup(\{A, C\}) = 50\%$
    - $conf(C \rightarrow A) = sup(\{A, C\})/sup(\{C\}) = 100\%$

Mining Association Rules

Problem Definition

- Given:
  - data set of transactions *D*
  - minimal support *minsup*
  - minimal confidence *minconf*
- Obtain:
  - **all** association rules

    $$X \rightarrow Y \quad (s = Sup, c = Conf)$$

    such that

    $$Sup \geq minsup \text{ and } Conf \geq minconf$$

Apriori Algorithm

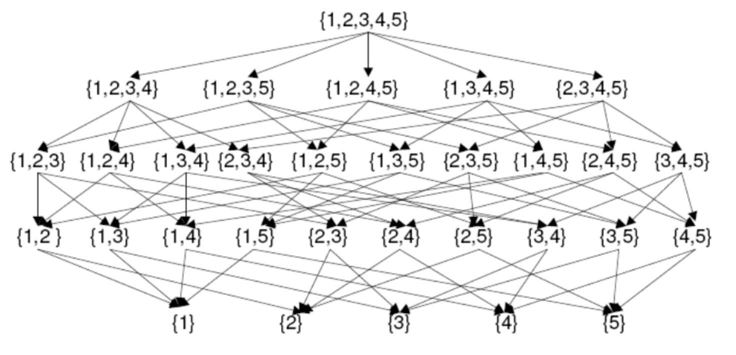
The **Apriori Algorithm** [Agrawal and Srikant, 1994] works in two steps:

1. **Frequent itemset generation**
   - itemsets with *support* $\geq$ *minsup*
2. **Rule generation**
   - generate all confident association rules from the frequent itemsets, i.e. rules with *confidence* $\geq$ *minconf*

Apriori Algorithm (cont.)

- Problem:
  - there is a very large number of candidate frequent itemsets!
    - for transactions with k items, there are $2^k - 1$ distinct subsets.
- Downward Closure Property
  - every subset of a frequent itemset must also be frequent.
    - ex: if $\{A1, A2, A4\}$ is frequent, so is $\{A1, A2\}$ because every transaction containing $\{A1, A2, A4\}$ also contains $\{A1, A2\}$.
  - thus, every superset of an infrequent itemset is also infrequent.
    - ex: if $\{A1, A2\}$ is infrequent, so is $\{A1, A2, A4\}$.
- Apriori Pruning Principle:
  - if an itemset is below the minimal support, discard all its supersets.

Search Space for 5 items



Example - 1 (cont.)

- Apriori enumerates and counts the support of patterns with increasing length.
- Starts looking for frequent itemsets of size 1 (F_1), assuming $minsup = 50\%$ (2 transactions)
- $C_1 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$

TID	ITEM-SET
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

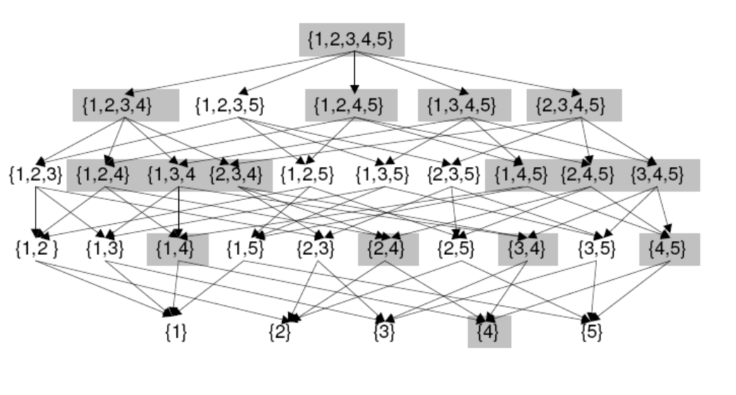
→

ITEM-SET	Support
{1}	2
{2}	3
{3}	3
{4}	1
{5}	3

- $F_1 = \{\{1\}, \{2\}, \{3\}, \{5\}\}$

Example - 1 (cont.)

- Filtered Search Space for 5 items (after removing item “4”)



Example - 1 (cont.)

- Looks for frequent itemsets of size 2 (F_2) from frequent itemsets of size 1 (F_1)
- Candidates $C_2 = \{\{a, b\} | \{a\} \in F_1 \wedge \{b\} \in F_1\}$
- $C_2 = \{\{1, 2\}, \{1, 3\}, \{1, 5\}, \{2, 3\}, \{2, 5\}, \{3, 5\}\}$

ITEM-SET	Support
{1,2}	1
{1,3}	2
{1,5}	1
{2,3}	2
{2,5}	3
{3,5}	2

- $F_2 = \{\{1, 3\}, \{2, 3\}, \{2, 5\}, \{3, 5\}\}$

Example - 1 (cont.)

- Looks for frequent itemsets of size 3 (F_3) from frequent itemsets of size 2 (F_2)
- Generation:
  $C0_3 = \{\{a, b, c\} | \{a, b\} \in F_2 \wedge \{a, c\} \in F_2\}$
- Filter:
  $C_3 = \{\{a, b, c\} | \{a, b, c\} \in C0_3 \wedge \forall x \in \{a, b, c\}\ S - \{x\} \in F_2\}$
- $C_3 = \{\{2, 3, 5\}\}$

ITEM-SET	Suporte
{2,3,5}	2

- $F_3 = \{\{2, 3, 5\}\}$
- There are no frequent itemsets of size 4

Step 1 - Identifying Frequent Itemsets

- Candidate generation (Self-Join step)
  - generates new candidate k-itemsets based on the frequent (k-1)-itemsets found in the previous iteration.
- Candidate pruning (Prune step)
  - eliminates some of the candidate k-itemsets using the support-based pruning strategy.

Step 1 - Identifying Frequent Itemsets (cont.)

- Self-Join Example:

  Given the size k candidates
  $\{A, B, C\}$
  $\{A, B, D\}$
  $\{A, C, D\}$
  $\{B, C, D\}$
  $\{A, B, E\}$
  $\{B, C, E\}$
  and assuming that in each itemset the items are lexicographically sorted

- Which are the candidates of size $k + 1$?

- What is the most efficient way of finding them (without repetitions)?

- Look for pairs of sets with the same prefix of size $k-1$
  $\{A, B, C\}$ and $\{A, B, D\}$
- Combine both, keeping the prefix
  $\{A, B, C, D\}$
- This way
  - No frequent set is unnoticed
  - No candidate is generated more than once

Step 1 - Identifying Frequent Itemsets (cont.)

- Prune Example:

  $F_3 = \{\{A, B, C\}, \{A, B, D\}, \{A, C, D\}, \{A, C, E\}, \{B, C, D\}\}$

  $C_4 = \{\{A, B, C, D\}, \{A, C, D, E\}\}$

  but $\{A, C, D, E\}$ can be pruned away

  because $\{A, D, E\} \notin F_3$

- Note:
  - Prune maintains the completeness of the process

Step 2 - Rule Generation

- Given a frequent set $\{A, B, C, D\}$
- Which are the possible rules?
  - $\{A, B, C\} \rightarrow \{D\}$
  - $\{A, B, D\} \rightarrow \{C\}$
  - $\{A, B\} \rightarrow \{C, D\}$
- How to generate them systematically?
- How to reduce the search space?

Step 2 - Rule Generation (cont.)

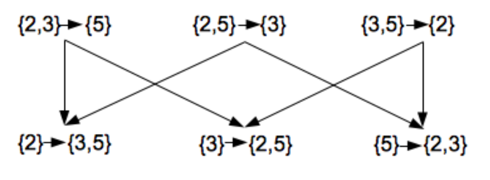
- The rules are generated as follows:
  - generates all non-empty subsets s of each frequent itemset l
  - for each subset s computes the confidence of the rule $(l - s) \rightarrow s$
  - selects the rules whose confidence is higher than *minconf*

Consider again

Cliente (TID)	Itens (Item-set)
100	1, 3, 4
200	2, 3, 5,
300	1, 2, 3, 5,
400	2, 5,

and $I = \{2,3,5\} (= F_3)$

- Rules generated from the frequent itemset $\{2,3,5\}$



- Select rules $(I - a) \rightarrow a$, where $a \subseteq I$, with *minconf* $= 1$

$$conf((I-a) \rightarrow a) = \frac{sup(I)}{sup(I-a)}$$

Step 2 - Rule Generation (cont.)

- Rules with 1 consequent

$\{2,3\} \rightarrow \{5\}$ (conf= 2/2)
$\{2,5\} \rightarrow \{3\}$ (conf= 2/3) eliminated because *minconf* $= 1$
$\{3,5\} \rightarrow \{2\}$ (conf= 2/2)

- Rules with 2 consequents

$\{3\} \rightarrow \{2,5\}$ (conf= 2/3) eliminated because *minconf* $= 1$

- we don't need to worry about rules with item 3 in the consequent, because any rule obtained from $\{2,5\} \rightarrow \{3\}$ will have a *conf* $< 2/3$

> Moving items from the antecedent to the consequent never changes support and never increases confidence.

Number of DB scans

- 1 to count frequencies of C_1
- C_2 built in memory
- 2 to count frequencies of C_2
- ...
- n to count frequencies of C_n
- Rule generation does not need to scan DB
- Number of scans is n
  - if the size of the largest frequent set is n or $n-1$

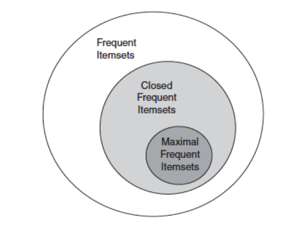
Complexity factors

- Number of items
- Number of transactions
- Minimal support
- Average size of transactions
- Number of frequent sets
- Average size of a frequent size
- Number of DB scans
  - k or $k+1$, where k is the size of the largest frequent set

Compact Representation of Itemsets

- The number of frequent itemsets produced from a transaction data set can be very large.
- It is useful to identify a small representative set of itemsets from which all other frequent itemsets can be derived.

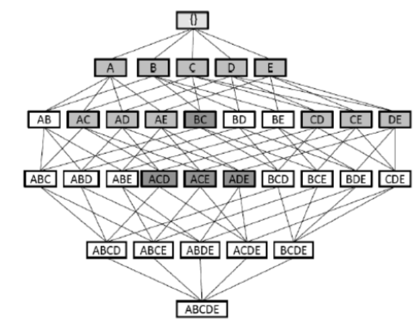
- Two such representations are:
  - maximal
  - closed



Compact Representation of Itemsets (cont.)

- s is a **closed frequent itemset** if it is a frequent itemset that has no frequent supersets with the same support.
- Example: find closed frequent itemsets with *minsup* $= 30\%$

TID	Itemset
1	A D E
2	B C D
3	A C E
4	A C D E
5	A E
6	A C D
7	B C
8	A C D E
9	B C E
10	A D E

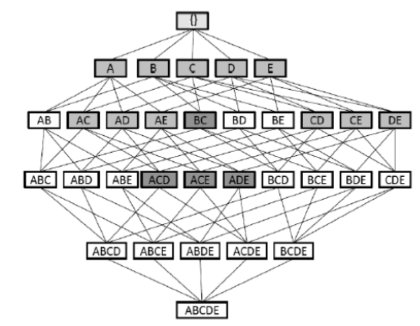


closed frequent itemsets are:
$\{A\}, \{C\}, \{D\}, \{E\}, \{A, C\}, \{A, D\}, \{A, E\},$
$\{B, C\}, \{C, D\}, \{C, E\}, \{A, C, D\}, \{A, C, E\}, \{A, D, E\}$

Compact Representation of Itemsets (cont.)

- s is a **maximal frequent itemset** if it is a frequent itemset for which none of its supersets is frequent.
- Example: find maximal frequent itemsets with $minsup = 30\%$

TID	Itemset
1	A D E
2	B C D
3	A C E
4	A C D E
5	A E
6	A C D
7	B C
8	A C D E
9	B C E
10	A D E



maximal frequent itemsets are:

$\{B, C\}, \{A, C, D\}, \{A, C, E\}, \{A, D, E\}$

Compact Representation of Itemsets (cont.)

- From the maximal itemsets is possible to derive all frequent itemsets (not their support) by computing all non-empty intersections.
  - subsets of the maximal frequent itemset $\{A, C, D\}$ are frequent itemsets
  - $\{A\}$, $\{C\}$, $\{D\}$, $\{A, C\}$, $\{A, D\}$, $\{C, D\}$
- The set of all closed itemsets preserves the knowledge about the support values of all frequent itemsets.
  - $\{D, E\}$ is a non closed frequent itemset. What is its support?
  - As it is not closed, its support must be equal to one of its immediate supersets.
  - look for the most frequent closed itemset that contains $\{D, E\}$: $\{A, D, E\}$
  - $sup(\{D, E\}) = sup(\{A, D, E\})$
- There are algorithms that take advantage of this compact representation of frequent itemsets.

Too many rules ...

- The association rule algorithms tend to generate an excessive number of rules (for some problems, there can be thousands).
- Too many rules leads to model's interpretability lack.
- How can we reduce this number?
  - Changing the parameters: *minsup*, *minconf*
  - Restrictions on items: which items are relevant?
  - Summarization techniques: can we represent subsets of rules by a single representative rule?
  - Filter rules: improvement, measures of interest, ...

How to measure the improvement of a rule?

Improvement [Bayardo and Ag, 2000]

- **Improvement** of a rule is the minimum difference between its confidence and the confidence of any of its immediate simplifications.

  $$improv(A \rightarrow C) = min(\{conf(A \rightarrow C) - conf(As \rightarrow C) \mid As \subseteq A\})$$

- Example:
  - $R_1 : \{eggs, flower, bread\} \rightarrow \{sugar\}(conf = 0.505)$
  - $R_2 : \{eggs, flower\} \rightarrow \{sugar\}(conf = 0.5)$
  - $improv(R_1)$ is at most 0.005
  - with a *minmprov* of 0.01, R_1 is excluded.

Are all the rules interesting?

- Are all the discovered patterns interesting?
- In recent years, several measures have been proposed to extract interesting patterns.
- The idea is to select a subset of rules, that somehow are more relevant.
- Interesting rule (Silberschatz & Tuzhilin,95)
  - Unexpected, surprising to the user
    - Measure of interest: deviation from the expected or from the initial belief
  - Useful, actionable
    - Measure of interest: estimated benefit

How to measure the interest of a rule?

- Subjective measures: based on user's belief in the data (ex: unexpectedness, novelty, actionability, confirm hypothesis user wishes to validate)
  - These measures are hard to incorporate in the pattern discovery task.
- Objective measures: based on facts, statistics and structures of patterns (ex: support and confidence), independent of the domain considered.
  - For instance, patterns that involve mutually independent items or cover very few transactions are considered uninteresting.

How to measure the interest of a rule? (cont.)

Typically

- $A \rightarrow B$ is interesting if A and B are not statistically independent
- if A and B are statistically independent, the occurrence of A does not affect the probability of occurrence of B

$$sup(A \cup B) \approx sup(A) * sup(B)$$

$$conf(A \rightarrow B) \approx conf(\emptyset \rightarrow B)$$

- $A \rightarrow B$ may have high support and confidence and still not be interesting.
  - $\{butter\} \rightarrow \{bread\}(sup = 5\%, conf = 95\%)$
  - it is not unexpected
  - it is not useful

How to measure the interest of a rule? (cont.)

- A measure of interest should evaluate the deviation from independence.
- A rule is unexpected as it deviates from independence.
- There are different approaches to measure this deviation:
  - *lift*
  - *conviction*
  - χ^2
  - *correlation*
  - ...

Measures of Interest: limitations of support and confidence

- Assume we are interested in studying the relationship between people who drink tea and coffee.
- We summarize the preferences of 1000 people

	Coffee	$\neg$*Coffee*	
Tea	150	50	200
$\neg$*Tea*	650	150	800
	800	200	1000

- How interesting is the rule $Tea \rightarrow Coffee$?
- $sup = 150/1000 = 15\%$ and $conf = 150/200 = 75\%$
- The confidence of the rule is high, however the likelihood of a person drinking coffee regardless of drinking tea is 80%.
- Knowing that a person drinks tea actually decreases the probability of drinking coffee (from 80% to 75%).
- Thus, the rule is indeed deceitful.
- High confidence rules can be misleading.

Measures of Interest: LIFT

- **lift** is the ratio between confidence of the rule and the support of the itemset appearing in the consequent:

$$lift(A \rightarrow B) = \frac{conf(A \rightarrow B)}{sup(B)} = \frac{sup(A \cup B)}{sup(A)sup(B)}$$

- Measures the influence of A in the presence of B.
- $lift = 1$: A and B are independent ($sup(A \cup B) = sup(A)sup(B)$).
- $lift < 1$: A and B are negatively correlated.
- $lift > 1$: A and B are positively correlated.

- $lift(Tea \rightarrow Coffee) = 0.15/(0.2 * 0.8) = 0.9375$
- negative correlation between tea and coffee drinkers.

Measures of Interest: LIFT (cont.)

- The **lift** is a measure of the deviation from a rule $A \rightarrow B$ regarding the statistical independence between the antecedent A and consequent B.

- Takes values between 0 and infinity:
  - a value close to 1 indicates that A and B often appear together
    - the occurrence of A has no effect on the occurrence of B.
  - a value smaller than 1 indicates that A and B appear less frequently than expected together
    - the occurrence of A has a negative effect on the occurrence of B, i.e. the occurrence of A is likely to lead to the absence of B.
  - a value greater than 1 indicates that A and B appear more often together than expected
    - the occurrence of A has a positive effect on the occurrence of B, i.e. the occurrence of A increases the likelihood of occurrence of B.

Measures of Interest: Conviction

- **lift** measures co-occurrence only (not implication) and is symmetric with respect to antecedent and consequent, i.e. $\mathit{lift}(A \rightarrow B) = \mathit{lift}(B \rightarrow A)$
- **conviction** is a measure proposed to tackle some of the weaknesses of *confidence* and **lift**.
- Unlike **lift**, **conviction** is sensitive to rule direction. It indicates the departure from independence of A and B taking into account the implication direction.
- Is inspired in the logical definition of implication and attempts to measure the degree of implication of a rule.

- **conviction** of a rule $A \rightarrow B$ is the ratio between
  - the expected frequency that A occurs without B, if A and B were independent
  - the observed frequency that the rule makes of incorrect predictions.
- Is the inverse **lift** of the rule $R' = A \rightarrow \neg B$.

$$conviction(A \rightarrow B) = \frac{1 - sup(B)}{1 - conf(A \rightarrow B)} = \frac{sup(A)sup(\neg B)}{sup(A \cup \neg B)}$$

Measures of Interest: Conviction (cont.)

- $conviction(A \to B) = 1$ indicates independence between A and B.
- A high value of **conviction** means that the consequent depends strongly on the antecedent.
- **conviction** increases a lot when *confidence* gets closer to 1.
- Example:
  - $sup(female) = 0.5$, $sup(mother) = 0.2$
  - $conf(mother \to female) = 1$
  - $lift(mother \to female) = 0.2/(0.2 * 0.5) = 2$
  - $conviction(mother \to female) = (1 - 0.5)/(1 - 1) = \infty$

Improving Apriori

- Challenges of Frequent Pattern Mining
  - Multiple scans of transaction database
  - Huge number of candidates
  - Tedious workload of support counting for candidates
- Improving Apriori: general ideas
  - Reduce number of transaction database scans
  - Shrink number of candidates (*bottleneck* of Apriori)
  - Facilitate support counting of candidates
- Some methods that improve Apriori's efficiency
  - Partitioning [Savasere et al., 1995]
  - Sampling [Toivonen, 1996]
  - Dynamic Itemset Counting [Brin et al., 1997]
  - Frequent Pattern Projection and Growth (FP-Growth) [Han et al., 2004]

Association Rules: Conclusions

- GOAL: Finding associations
- Association rule mining:
  - Frequent itemsets (requires min support)
  - Association rules (requires min confidence)
    - Probabilistic implications
- One of the most used data mining tools
  - Problem: generates too much rules
  - Pattern compression and pattern selection
- Several algorithms:
  - Apriori is the most known algorithm
  - There are variants of Apriori that return exactly the same patterns!
  - Completeness: find all rules.

References

References

Aggarwal, C. C. (2015).
Data Mining, The Texbook.
Springer.

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., and Verkamo, A. I. (1996).
Fast discovery of association rules.
In *Advances in Knowledge Discovery and Data Mining*, pages 307–328. American Association for Artificial Intelligence.

Agrawal, R. and Srikant, R. (1994).
Fast algorithms for mining association rules in large databases.
In *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB '94, pages 487–499. Morgan Kaufmann Publishers Inc.

Brin, S., Motwani, R., Ullman, J. D., and Tsur, S. (1997).
Dynamic itemset counting and implication rules for market basket data.
In *Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data*, volume 26, pages 255–264. ACM.

Domingo, C., Gavalda, R., and Watanabe, O. (1998).
On-line sampling methods for discovering association rules.

References (cont.)

Gama, J. (2016).
Association rules.
Slides.

Gama, J., Oliveira, M., Lorena, A. C., Faceli, K., and de Leon Carvalho, A. P. (2015).
***Extração de Conhecimento de Dados - Data Mining*.**
Edições Sílabo, 2nd edition.

Han, J., Kamber, M., and Pei, J. (2011).
***Data Mining: Concepts and Techniques*.**
Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition.

Han, J., Pei, J., Yin, Y., and Mao, R. (2004).
Mining frequent patterns without candidate generation: A frequent-pattern tree approach.
Data Mining and Knowledge Discovery, 8(1):53–87.

Jorge, A. (2016).
Association rules.
Slides.

Liu, B. (2011).
***Web Data Mining. Exploring Hyperlinks, Contents, and Usage Data*.**
Springer, 2nd edition.

References (cont.)

Savasere, A., Omiecinski, E., and Navathe, S. B. (1995).
An efficient algorithm for mining association rules in large databases.
In *Proceedings of the 21th International Conference on Very Large Data Bases*, VLDB '95, pages 432–444. Morgan Kaufmann Publishers Inc.

Tan, P.-N., Steinbach, M., and Kumar, V. (2005).
***Introduction to Data Mining*.**
Addison Wesley.

Toivonen, H. (1996).
Sampling large databases for association rules.
In *Proceedings of the 22th International Conference on Very Large Data Bases*, VLDB '96, pages 134–145. Morgan Kaufmann Publishers Inc.

Torgo, L. (2017).
***Data Mining with R: Learning with Case Studies*.**
Chapman and Hall/CRC, 2nd edition.